Est. 1841 | YORK ST JOHN UNIVERSITY

Panesar, Kulvinder ORCID: https://orcid.org/0000-0002-4523-7218 (2018) HOW CAN ONE EVALUATE A CONVERSATIONAL SOFTWARE AGENT FRAMEWORK? In: 7th International Conference on Meaning and Knowledge Representation, 4-6 July 2018, ITB (Institute of Technology Blanchardtown), Dublin.

Downloaded from: http://ray.yorksj.ac.uk/id/eprint/3402/

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. Institutional Repository Policy Statement

RaY

Research at the University of York St John

For more information please contact RaY at ray@yorksj.ac.uk

7th INTERNATIONAL CONFERENCE ON MEANING AND KNOWLEDGE REPRESENTATION (4, 5 and 6 July, 2018)

Session 8 – 6th July 2018

How can one evaluate a conversational software agent framework?

Dr Kulvinder Panesar
Senior Lecturer in Computer Science
York St John University

Overview

- Context – conversational agent space
- Linguistic approach
- Building blocks of a Linguistic CSA
- Drivers of LING-CSA
- Conceptual framework LING-CSA
- Implementation of LING-CSA
- How to evaluate LING-CSA?
- Evaluations and findings
- Evaluation – conclusions & recommendation

1 Conversational agents space

Figure 1 – Types of conversation
Source: FactionXYZ

Figure 2 – Hype cycle and myths
Source: FactionXYZ

2. Linguistic Approach

Conversational Software Agents (CSA)

- Challenges of NLU and meaning
- NL -> functional system
- Periñán-Pascual (2013):eligibility
- Communication-cognition
- Approach – unique framework, model/theory interaction, communicative
- Language levels, interface between syntax, semantic, and pragmatics
- Language Model: RRG and the clause

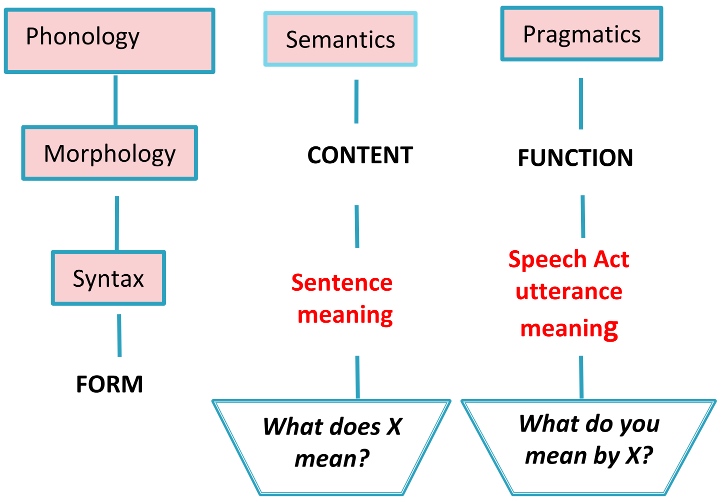


Figure 1: Language interfaces

- Simple sentences ->Linguistic act (Speech Act) – SA
- Understand the utterance
- Agent attributes
- (Utterance) Message from USER → AGENT
- Agent's belief – Knowledge representation (KR)
- Plan-based dialogue (response) Message AGENT → USER

3. Building blocks of a CSA

AGENT	+	INTELLIGENT DIMENSION (S)	=	INTELLIGENT AGENT

INTELLIGENT TAXONOMY Behavioural, Social, Ambient, Collective, Genetic, and **COGNITION**

COGNITION = BDI + Rational Interaction

CA = Interpretation + Dialogue Mgt + Response Generator

CSA = CA +RRG +SA + COGNITIVE + KB (Panesar,2017)

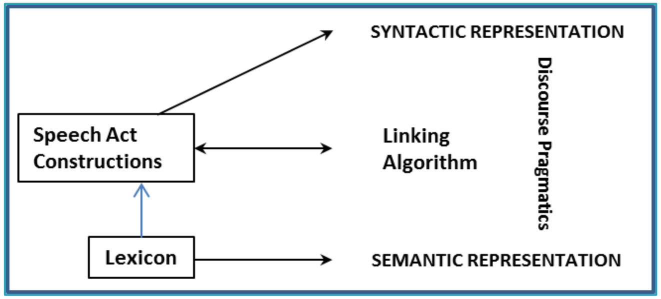


Figure 3– Reorganisation of RRG

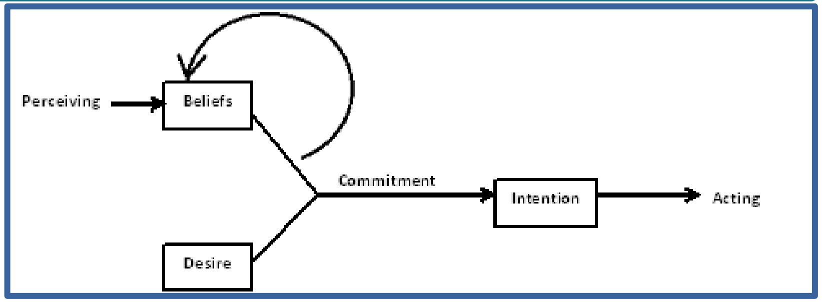


Figure 4– A BDI model of an intelligent agent (Allen, 1995)

4.Drivers of LING-CSA

- Investigate the *integration, intersection and interface* of the language, knowledge, and speech act constructions (SAC) based on a grammatical object (Nolan, 2014), and the sub-model of belief, desires and intention (BDI) (Rao and Georgeff, 1995) and dialogue management (DM) for natural language processing (NLP).
- A long-standing issue within NLP CSA systems is *refining the accuracy of the interpretation of meaning* to provide a realistic dialogue to support the human-to-computer communication.

5. Conceptual Framework: LING-CSA

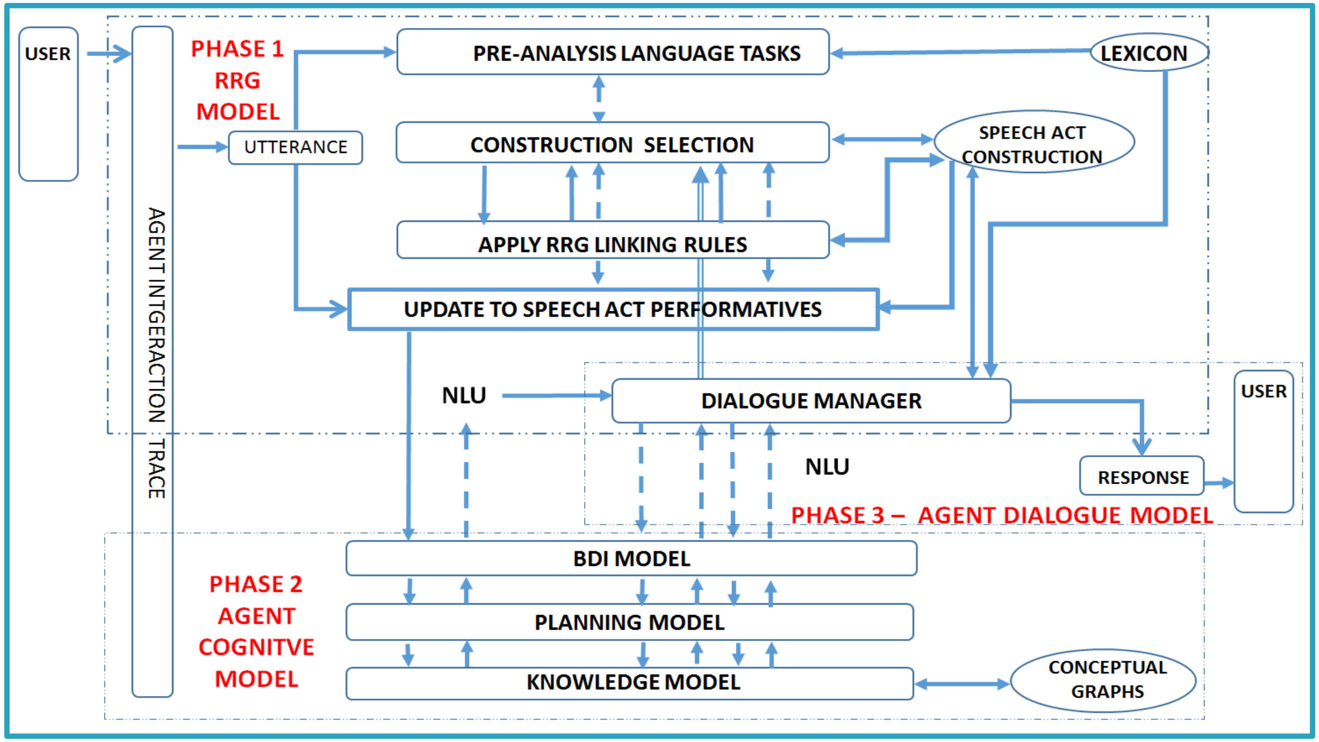


Figure 5- Conceptual framework of the Conversational Software Agent (Panesar, 2017)

PHASE 1 – Role and Reference Grammar (RRG) Language Model
PHASE 2 – Agent Cognitive Model interfaces with:
BDI Model, Planning Model, Knowledge Model
PHASE 3 – Agent Dialogue Model (Dialogue Mgnt > RRG Model

6. Phase 2 – Agent Cognitive Model Design Framework

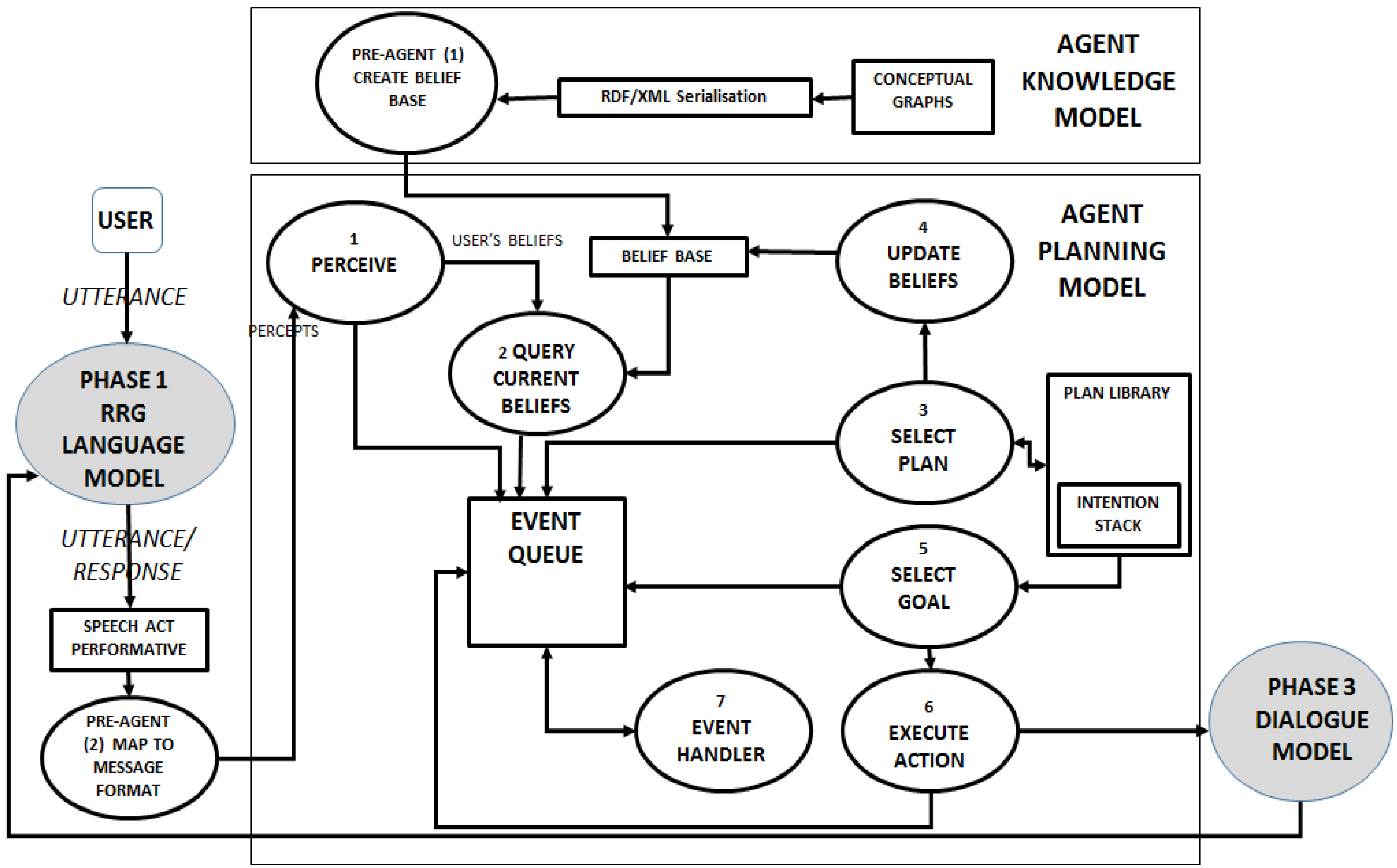


Figure 6 – The Agent Cognitive Model – Design Framework (Panesar, 2017)

7.How to evaluate LING-CSA?

- ISO 9241 concept of usability?
- Turing test?
- Define goals
  - Phase model based
  - Interfacing, intersection and integration
- Customised multi-approach assessment
- Testing strategy
  - Grammatical testing and NLP tasks
  - Software engineering
  - Knowledge representation
  - Agent practice and environment
  - RRG specific tests and goals of linguistic theory
- Evaluation criteria (goal-driven)

8. Evaluation criteria

Phase 1 – RRG Language Model

- Criteria 1– Could the system present a mapping of the syntactic representation to a semantic representation, for the utterance taking the form of a simple sentence?
- Criteria 2 – Could the system present an adequate explanation of the NLU of the utterance?
- Criteria 3 – Could the system demonstrate the SAC use in the manipulation of the utterance?
- Criteria 4 – Can the dialogue manager interface the language model?

Phase 2 and 3 – Agent Cognitive Model and Agent Dialogue Model

- Criteria 5 – Could the system demonstrate the agent BDI and knowledge representation?
- Criteria 6 – Could the system represent the user's BDI states?
- Criteria 7 – Could the system query the knowledge base for a fact (from the speech act performative)
- Criteria 8 – Could the system devise an appropriate plan based on the BDI states?
- Criteria 9 – Could the system generate a grammatically correct response in RRG based on the agent's knowledge?

9. Evaluation Phase 1 – RRG :LSC and LS

- RRG is a functional model. It views language as a communicative social action.
- Layered structure of the clause (LSC) = PREDICATE + ARGUMENT + NON-ARGUMENTS.
- Logical Structure (LS) – semantic meaning of the sentence.
- Lexicon - mental dictionary - lexical entries contain semantic features and constraints.
- It maps the syntax(structure): LSC ⟺ semantic (meaning): LS the actual form of the sentence using two different LINKING ALGORITHMNS.
- RRG parser (algorithm) checks the grammar (rules) of English. Specialised parser (CSA)
- RRG facilitates syntactic, semantic and information structure (FOCUS & TOPIC)

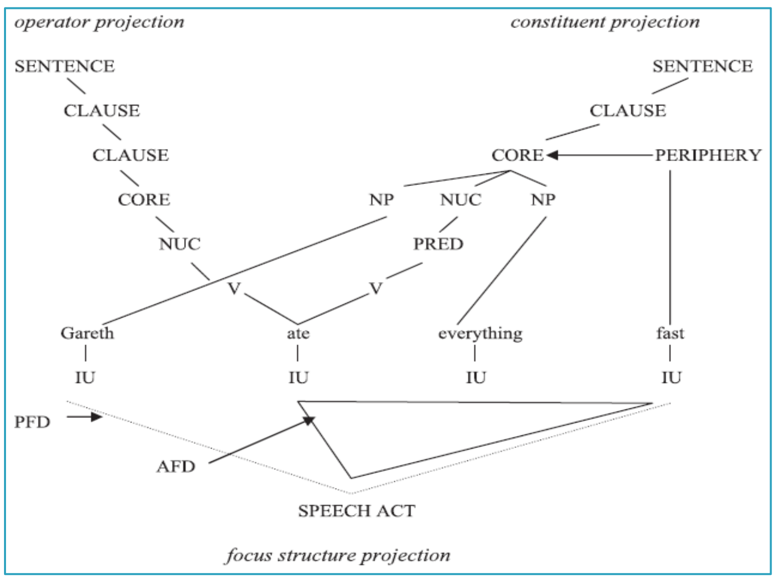


Figure 7 - An English sentence with three representations

RRG steps (Panesar, 2017)

Worked example

Gareth ate everything fast

(BNC ADY 1079) (Butler et al, 2009) –> Figure 5

SYNTACTIC:
SENTENCE (CLAUSE (<CORE> <NP> gareth (<NUC> (<PRED> <V> ate)) (<NP> (everything))) (PERIPHERY fast)

SEMANTIC:
[<IF> ASS <TNS> PST, do'(ACT:Gareth, (eat'(Gareth <NOM>, pizza <ACC>)])] & INGR consumed' (UND:pizza)]

10. Evaluation Phase 1 – RRG : Data sources – Lexicon, SA, & SAC

LEXICAL ENTRY	POS-TYPE	VERB TENSE/ ASPECT	DEF	P TYPE	NO	GR	CASE	ANIM	HUM	LOGICAL STRUCTURE (LS)
ate	VERB	PST	DEF+/-	3	SG	M/F	DNA	ANIM	HUM	<tns:pst <do'(x, [eat'(x, y)]) & BECOME consumed' (y) >>
eat	VERB	PRS/ FUT	DEF+/-	3	SG	M/F	DNA	ANIM	HUM	<tns:prs <do'(x, [eat'(x,y)]) & BECOME consumed'(y)] >> <tns:fut <do'(x, [eat'(x,y)]) & BECOME consumed'(y) >>
eating	VN	PROG	DEF+/-	3	SG	M/F	DNA	ANIM	HUM	<tns:prs <asp:prog <do'(x, [eat'(x, y)]) & BECOME consumed' (y)] >>>
is	VBE	DNA	DEF+	DNA	DNA	DNA	DNA	DNA	DNA	be'(x,[pred'])
hungry	ADJ	DNA	DNA	DNA	DNA	M/F	DNA	ANIM	HUM	DNA
restaurant	N	DNA	DEF+/-	DNA	SG/PL	DNA	DNA	DNA	DNA	DNA

Table 1 – Snapshot of the Lexicon (Panesar, 2017)

Speech (linguistic) Act (SA) Theory (Searle, 1969) – message types as in Figure 6

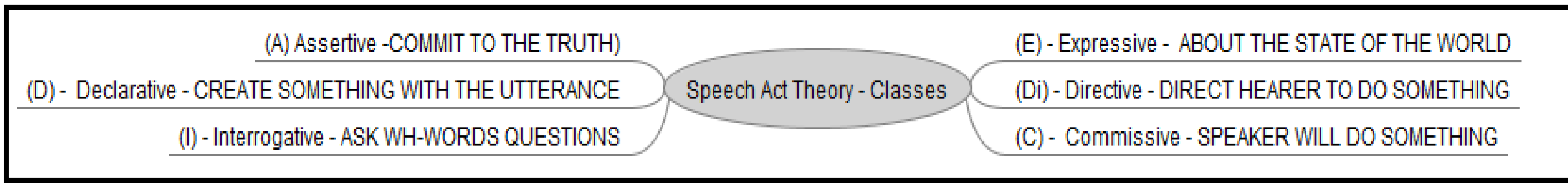


Figure 8– Speech Act message types

3 actions associated with an utterance include:

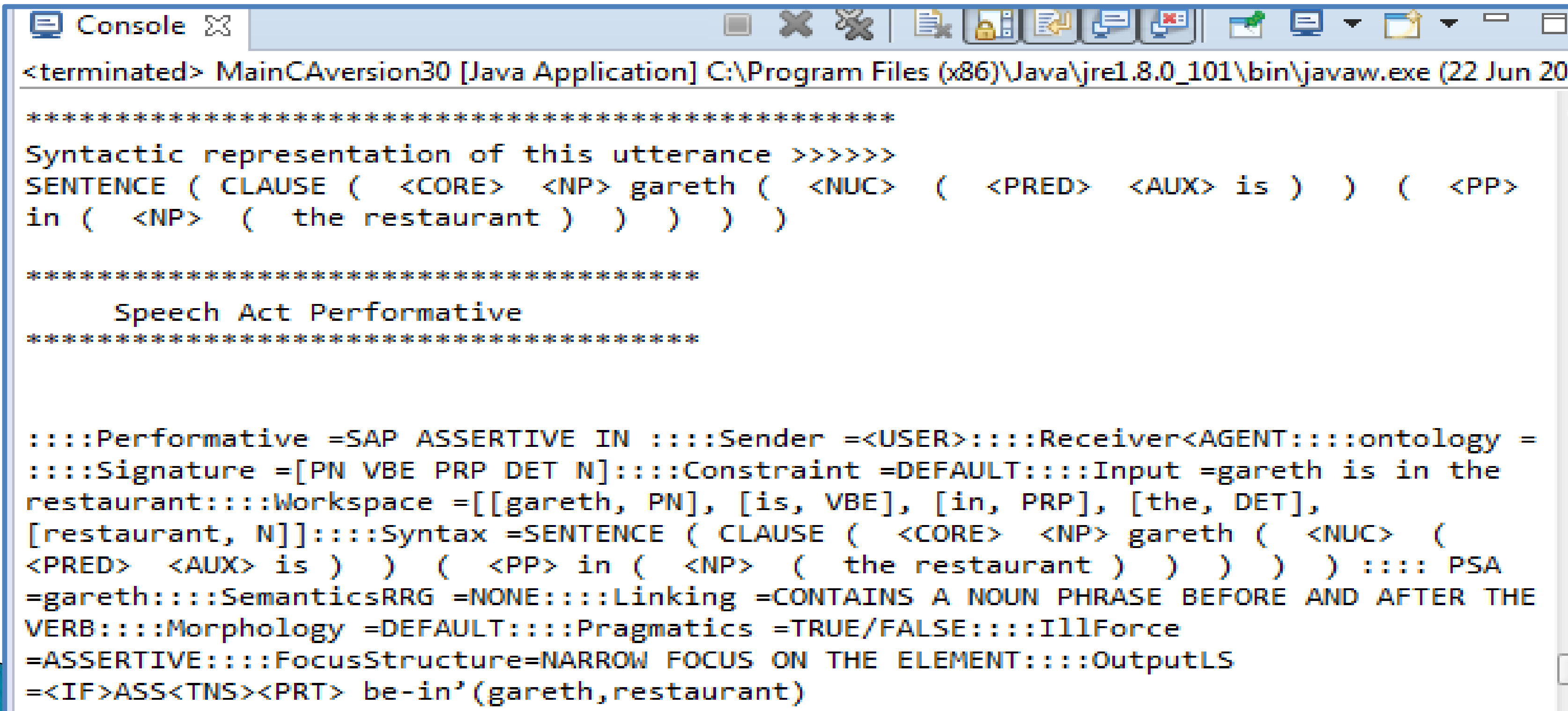
- Locution:
- Illocution: illocutionary act (speaker's intention) [SI] for A, Di and I message types
- Perlocution:

Figure 9 – Empty SAC (Speech Act Construction) (Panesar, 2017)

ASSERTIVE:ATE RRG [NP VERB NP], [PN VERB], [ADV PN VERB DET N], [PN VERB N ADJ], [PRP DET N PN VERB DET N], [PN VBE VERB N], [PN PRP DET N PRP DET N], [PRO VERB DET N], [PN VERB NP], [PN VERB DET N], [NP VERB QNT N], [DET N VERB DET N], [DET N VERB QNT N], [NP VERB (DET) (ADJ) N (ADJ)], [PN VERB DET N ADJ], [PN VERB (DET) ADV N ADJ], [PN VERB DET N PRP DET N], [PN, VERB, N, PRP, DET, N], [PN VERB N PRP DET N] RRG NONE RRG UTTINPUT RRG WKSPACE RRG DEFAULT ASSUMPTION (1ST NP = 'ACTOR") RRG NO PARTICULAR SPEC RRG NONE RRG CONTAINS A NOUN PHRASE BEFORE AND AFTER THE VERB RRG DEFAULT RRG TRUE/FALSE RRG ASSERTIVE RRG NARROW FOCUS ON THE ELEMENT RRG LOG STRUCTURE TO ADD

11. Evaluations (Phase 1 – RRG Model)

- Aim- assessment of Criteria 1 -4
- Each specific construal (either an utterance or response) –two steps.
  1. Find the matching SA construction of that specific predicating element. In Figure 2: ‘*is*’ and selected SAC of assertive.
  2. Select the matching signature pattern -> **[PN, VBE, PRP, DET, N]**
- Updates > SAC first and extended SAP (Panesar, 2017)



```
Console
<terminated> MainCAversion30 [Java Application] C:\Program Files (x86)\Java\jre1.8.0_101\bin\javaw.exe (22 Jun 201
*************************************************
Syntactic representation of this utterance >>>>>>
SENTENCE ( CLAUSE (  <CORE>  <NP> gareth (  <NUC>  (  <PRED>  <AUX> is )  )  (  <PP>
in (  <NP>  (  the restaurant )  )  )  )  )

**************************************
     Speech Act Performative
**************************************


::::Performative =SAP ASSERTIVE IN ::::Sender =<USER>::::Receiver<AGENT::::ontology =
::::Signature =[PN VBE PRP DET N]::::Constraint =DEFAULT::::Input =gareth is in the
restaurant::::Workspace =[[gareth, PN], [is, VBE], [in, PRP], [the, DET],
[restaurant, N]]::::Syntax =SENTENCE ( CLAUSE (  <CORE>  <NP> gareth (  <NUC>  (
<PRED>  <AUX> is )  )  (  <PP> in (  <NP>  (  the restaurant )  )  )  )  ) :::: PSA
=gareth::::SemanticsRRG =NONE::::Linking =CONTAINS A NOUN PHRASE BEFORE AND AFTER THE
VERB::::Morphology =DEFAULT::::Pragmatics =TRUE/FALSE::::IllForce
=ASSERTIVE::::FocusStructure=NARROW FOCUS ON THE ELEMENT::::OutputLS
=<IF>ASS<TNS><PRT> be-in'(gareth,restaurant)
```

Figure 10 – Snapshot output of LING-CSA (Panesar, 2017)

12. Phase 1 – RRG & Speech Act Performative

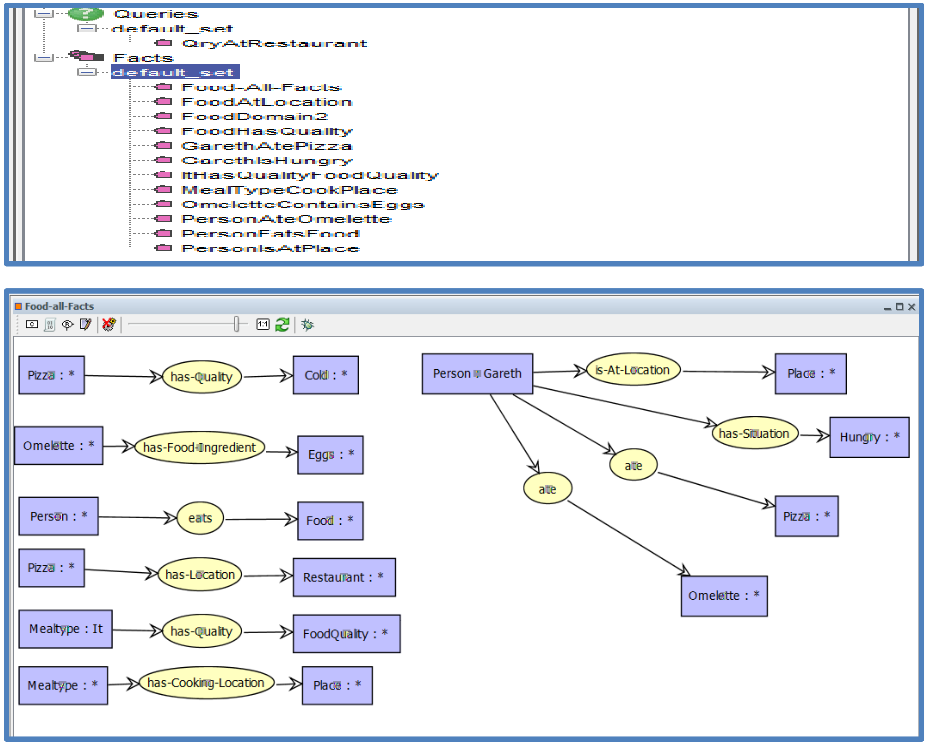
Based on the SAC with four additional attributes. Input to Phase 2-ACM

PERFORMATIVE: <ASSERTIVE:ATE>
:SENDER <USER>
:RECEIVER <AGENT-1>
:ONTOLOGY <FoodAndCookKB>
:CONTENT <do'(Gareth, (eat'(Gareth, pizza)])] & INGR consumed' (pizza)] everything>
SIGNATURE: [PN V NP ADJ] CONSTRAINT: Default
INPUT: Gareth ate everything fast
WORKSPACE: (Gareth, PN), (ate, VERB), (everything N), (fast, ADJ)
SEMANTICS: Contains a noun phase before and after the verb
CONSTRUCTION BODY SYNTAX: SENTENCE (CLAUSE (<CORE> <NP> gareth (<NUC> (<PRED> <V> ate)) (<NP> (everything))) (PERIPHERY fast) PSA: gareth SEMANTICS Linking: MORPHOLOGY:Default PRAGMATICS Illocutionary force: ASSERTIVE Focus structure: narrow focus on the element
OUTPUT [LS]: [<IF> ASS <TNS> PST, do'(ACT:Gareth, (eat'(Gareth <NOM>, pizza <ACC>)])] & INGR consumed' (UND:pizza)]

Table 2-Speech Act Construction Performative "ate" used as a message to the Agent Environment (Panesar, 2017)

13. Phase 2 – Agent Cognitive Model & Knowledge Model (Panesar, 2017)

SHARED and INDIVIDUAL BELIEFS cognitively → mental knowledge.



- **Conceptual graphs** (CGs) (Sowa,1986), Vocabulary, First order logic (FOL) created in COGUI as in Figure 7 and 8
- **Serialised** into RDF/XML (W3C SW), mapped to RDF Triple Stores – forms the agent's belief base – 446 lines (Table 3)
- KB ready for querying to check truth of the agent's beliefs
- Key Performance Indicators – representational and inferential adequacy

Figure 11 & 12– COGUI– Original KB of facts – graphically

Table 3 – Extract of a RDF triple Stores KB

No	Subject	Predicate	Object
1	http://www.lirmm.fr/cogui#ct_ad452f18-e654-4ae6-b3a1-b7320616283b	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2000/01/rdf-schema#Class
2	http://www.lirmm.fr/cogui#ct_fdc6d7d0-1314-4fb7-8428-51e122953250	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2000/01/rdf-schema#Class

14. Lexical Bridging Solution (Panesar,2017)

Reduce this semantic gap, by "building a lexical bridge (LB)" between the NL semantic and ontology semantics, with an aim to capture more of the meaning, by attempting to 'lexicalize the ontology'.

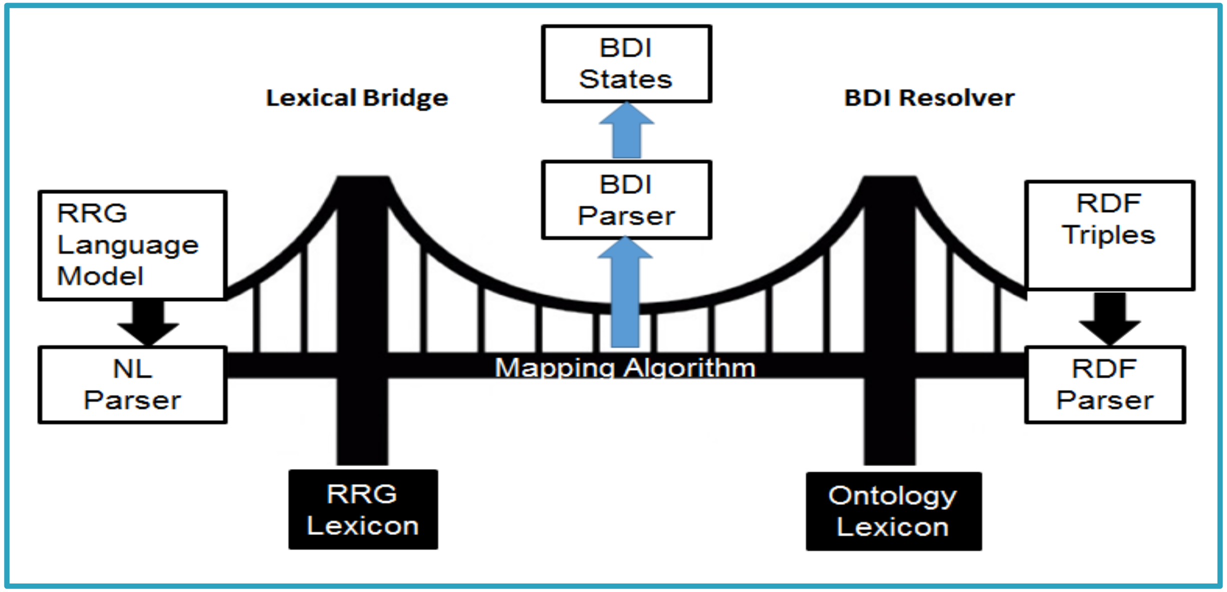


Figure 13 - Lexical Bridge for the CSA's belief base + BDI Parser to resolve the agent's BDI states

15. Evaluations, Findings, and Recommendations

Implementation outcomes :

- Proof-of-concept achieved;
- Dialogue Manager is common to Phase 1 and Phase 3

Findings

- RRG is fit for purpose ->linguistic engine for the CSA;
- RRG explains, describes linguistic phenomena; facilitates language processing and knowledge of language -> computationally adequate (Panesar, 2017)

Phase 1 – RRG Model Improvements:

1. All pronoun resolutions (E.g. 'Your', 'she', it' etc.)
2. Application of the propositional stranding rules
3. Complex sentences (extension of the RRG linking system)
4. Multi-lingual (additional lexicons) such as Spanish
5. Other SA classes such as emotive and commissives E.g analyse tweets
6. Include superlative adjectives/adverbs in the RRG Lexicon (E.g. 'spicier')
7. Invoke WordNet API for synonymous entries to the RRG Lexicon – ⇧value

Phase 2 Agent Cognitive Model working – 70% achieved Dialogue mgnt

Technical Challenge – Querying a natural language (NL) text against a knowledge representation (KR) of RDF triples poses a significant semantic gap

Conceptual solution (lexical bridge, BDI parser and RDF parser) (Panesar, 2017)

Single agent to multi-agent environment – an extended design framework

Content creation – via machine learning algorithms

References

- Allen, J. (1995) *Natural Language Understanding (2nd Ed.).* Benjamin-Cummings Publishing Co., Inc.
- Butler, C. S., and others (2008) *Layering in structural-functional grammars*. Linguistics, *46*(4), pp. 689-756.
- Cohen, P. R. and Levesque, H. J. (1990) *Intention Is Choice with Commitment*. Artificial Intelligence, 42(2), pp.213-261.
- Gartner. 2015. *Market Trends: Voice as a UI on Consumer Devices — What Do Users Want?* [Online]. Available: https://www.gartner.com/doc/3021226/market-trends-voice-ui-consumer [Accessed December, 2017].
- Nolan, B. (2013) *Constructions as Grammatical Objects : A Case Study of Prepositional Ditransitive Construction in Modern Irish.* In: Nolan, B. and Diedrichsen, E. (Eds.): Linking Constructions into Functional Linguistics: The Role of Constructions in Grammar. Amsterdam/Philadelphia, John Benjamins Publishing Company, pp.143-178
- Nolan, B. (2014) *Extending a Lexicalist Functional Grammar through Speech Acts, Constructions and Conversational Software Agents*. In: Nolan, B., & Periñán-Pascual, C. (Eds.): Language Processing and Grammars: The role of functionally oriented computational models. Vol.150. John Benjamins Publishing Company, pp.143-163.
- Panesar, K. (2017). *A linguistically centred text-based conversational software agent.* Unpublished PhD Thesis. School of Computing, Creative Technologies and Engineering. Leeds, UK, Leeds Beckett University.
- Periñán-Pascual, C. (2013) *A Knowledge-Engineering Approach to the Cognitive Categorization of Lexical Meaning*. VIAL-VIGO INTERNATIONAL JOURNAL OF APPLIED LINGUISTICS, 10, pp.85-104.
- Rao, A. S. and Georgeff, M. P. (1995). *BDI Agents: From Theory to Practice.* Paper presented at the ICMAS.
- Searle, J. R. (1969) *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.
- Sowa, J. F. and Way, E. C. (1986*) Implementing a Semantic Interpreter Using Conceptual Graphs.* IBM Journal of Research and Development, 30(1), pp.57-69.
- Van Valin, R. D. (2005) *Exploring the Syntax-Semantics Interface*. CUP..
- Pokahr, A., Braubach, L., Haubeck, C. and Ladiges, J. (2014) *Programming BDI Agents with Pure Java.* In: Multiagent System Technologies. Springer, pp.216-233
- Wallace, R. 2018. *Chatbots - a personal perspective*. Society for the study of Artificial Intelligence and Simulation of Behaviour (AISB) quarterly, 6.
- Visualistan, 2017. Chatbots: The Good, The Bad And The Ugly [Online]. Available: https://www.visualistan.com/2017/10/chatbots-good-bad-and-ugly-infographic.html [Accessed Feb 2018]
- Wooldridge, M. (2013) *Intelligent Agents*. In: Weiss, G. (Ed.): Multiagent Systems. USA, Massachusetts Institute of Technology, pp.3-50.

Thank you for listening!